REMARKS

This amendment is submitted in response to the outstanding Office Action, dated August 27, 2002, and is accompanied by a petition and fee for extension of time (one month). Claims 1 through 35 are presently pending in the above-identified patent application. No additional fee is due.

5   In the Office Action, the Examiner objected to the Specification for having an Abstract that exceeds 150 words and for a number of additional informalities. The Examiner objected to the Drawings for a number of informalities. The Examiner rejected Claims 1-5, 8, 10-14, 16-19, 21-26 and 28-35 under 35 U.S.C. § 102(b) as being anticipated by Chen et al., "Speaker, Environment and Channel Change Detection and Cluster via the Bayesian Information Criterion," Proc. of the DARPA Broadcast

10   News Workshop (Feb. 1998), hereinafter, referred to as "Chen." In addition, Claims 6 and 7 were rejected under 35 U.S.C. § 103(a) as being unpatentable over Chen in view of well known prior art.

The present invention automatically identifies speakers in an audio source by concurrently segmenting the audio source and clustering the segments corresponding to the same speaker.

15   FORMAL OBJECTIONS

In the Office Action, the Examiner objected to the Specification for having an Abstract that exceeds 150 words. The Abstract has been amended to ensure that it does not exceed 150 words.

The Examiner objected to the Specification for a number of additional informalities. In particular, the Examiner objected to a typographical error in the Abstract, at page 24, line 14. This

20   typographical error has been amended with the Abstract.

The Examiner objected to the Specification at page 3, as lacking an equation label. Applicants submit that the Equation on page 3 is part of the Summary, and is repeated again on page 9. An equation label is not required.

The Examiner objected to a number of terms and variables that are commonly used in

25   the BIC literature. For example, the Examiner objected to the terms "homogeneous segments" and "a single full covariance Gaussian" as being not clearly defined or explained. In addition, the Examiner objected to a number of symbols or parameters as lacking explanation or antecedent reference. Applicants submit that each of these terms, symbols or parameters are very well understood by a person of ordinary skill in the art. As evidence of such common understanding, the Examiner is pointed to a

30   number of references that have been cited with the present Office Action, including Chen et al., "Speaker, Environment and Channel Change Detection and Cluster via the Bayesian Information Criterion," Proc. of the DARPA Broadcast News Workshop (Feb. 1998); Chen, "Clustering Via the

Bayesian Information Criterion with Applications in Speech Recognition," IEEE Conf. on Acoustics, Speech and Signal Processing, 645-48 (1998); and United States Patent Numbers 6,421,645 and 6,424,946.

5   Thus, Applicants submit that the Specification and Abstract, as amended, are in full compliance of the patent rules and respectfully request that the objections be withdrawn.

The Examiner also objected to the Drawings as including a title in the flow charts and a number that was not underlined. Applicants submit formal drawings herewith that are believed to be in full compliance with the patent rules. Applicants submit that titles in a flow chart are permitted by the rules, and none of the numbers have been underlined. Thus, Applicants submit that the Drawings are in

10   full compliance of the patent rules and respectfully request that the objections be withdrawn.

PRIOR ART REJECTIONS

The Examiner rejected Claims 1-5, 8, 10-14, 16-19, 21-26 and 28-35 under 35 U.S.C. §102(b) as being anticipated by Chen and Claims 6 and 7 under 35 U.S.C. §103(a) as being unpatentable over Chen in view of well known prior art.

15   With regard to claim 1, the Examiner asserts that Chen discloses speaker, environment and channel change detection and clustering via the Bayesian Information Criterion for segmenting the audio stream into homogeneous regions according to speaker identity, environmental condition and channel condition and clustering speech segments into homogeneous clusters according to speaker identity, environmental condition and channel (citing page 1, paragraph 2) which reads on the claimed

20   "method of tracking a speaker in an audio source, said method comprising the steps of identifying potential segment boundaries in said audio source; and clustering homogeneous segments from said audio source substantially concurrently with said identifying step."

Applicants submit that while Chen does disclose segmenting an audio stream into homogeneous regions and clustering speech segments into homogeneous clusters, the audio stream is

25   **first** segmented and **then** clustered. A full reading of Chen indicates that Chen assumes that the process starts with a segmented audio stream. In particular, Section 3.3, second paragraph, indicates that the segmented data is provided by a third party, namely, the National Institute of Standards and Technology (NIST). Thus, it would be impossible for the "clustering of homogeneous segments from said audio source" to occur in Chen "substantially concurrently with said identifying step," as required

30   by each of the independent claims of the present invention.

As further evidence that the clustering in Chen is performed only after the audio stream has been segmented, Section 4.1 indicates that **each** segment is compared to **all** other segments before

clustering is finalized. In addition, Section 4.2, first paragraph indicates that the data set consists of an audio file that has been "hand-segmented into 824 short segments."

Thus, Chen does not disclose or suggest a "method of tracking a speaker in an audio source, said method comprising the steps of identifying potential segment boundaries in said audio source; and clustering homogeneous segments from said audio source **substantially concurrently** with said identifying step," as required by independent claims 1, 16, 30, 31, 32 and 33 of the present invention. Similarly, independent claims 23, 34 and 35 require that the segmentation and clustering are performed on the **"same pass"** through said audio source.

Dependent Claims 2 through 15, 17 through 22 and 24 through 29 were rejected under 35 U.S.C. §§102 or 103 as being unpatentable over Chen, alone or in combination with well known prior art. Claims 2 through 15, 17 through 22 and 24 through 29 are dependent on Claims 1, 16 or 23, respectively, and are therefore patentably distinguished over Chen, alone or in combination with well known prior art. because of their dependency from amended independent Claims 1, 16 or 23 for the reasons set forth above, as well as other elements these claims adds in combination to their base claim

In view of the foregoing, the invention, as claimed in Claims 1 through 35, cannot be said to be either taught or suggested by Chen, alone or in combination with well known prior art. Accordingly, applicants respectfully request that the rejection of claims 1 through 35 under 35 U.S.C. §§ 102 or 103 be withdrawn.

All of the pending claims, i.e., claims 1 through 35, are in condition for allowance and such favorable action is earnestly solicited.

If any outstanding issues remain, or if the Examiner has any further suggestions for expediting allowance of this application, the Examiner is invited to contact the undersigned at the telephone number indicated below.

The Examiner's attention to this matter is appreciated.

Respectfully submitted,

Kevin M. Mason

Dated: December 26, 2002

Kevin M. Mason
Attorney for Applicant(s)
Reg. No. 36,597
Ryan, Mason & Lewis, LLP
1300 Post Road, Suite 205
Fairfield, CT 06430
(203) 255-6560

VERSION MARKED TO SHOW ALL CHANGES

IN THE ABSTRACT:

5   Please amend the Abstract as follows:

[A method and apparatus are disclosed for] <u>Speakers are</u> automatically [identifying speakers from] <u>identified in</u> an audio (or video) source. The audio information is processed to identify potential segment boundaries[, corresponding to a speaker change]. [Thereafter, h]<u>H</u>omogeneous

10   segments [(generally corresponding to the same speaker)] are clustered <u>substantially concurrently with the segmentation routine,</u> and a cluster identifier is assigned to each identified segment. A segmentation subroutine identifies potential segment boundaries using the BIC model selection criterion. [A window selection scheme considers a relatively small amount of data in areas where new boundaries are very likely to occur, and the window size is increased when boundaries are not very

15   likely to occur. The window size increases in a slow manner when the window is small, and increases in a faster manner when the window gets bigger. When a segment boundary is found in a window, the next window begins after the detected boundary, using the minimal window size. BIC tests can be eliminated when they correspond to locations where the detection of a boundary is very unlikely.] A clustering subroutine uses a BIC model selection criterion to assign a cluster identifier to each of the

20   identified segments. If the difference of BIC values for each model [(BIC = BIC_1 - BIC_2)] is positive, the two clusters are merged. [The online clustering technique of the present invention involves the K clusters found in the previous iterations (or calls to the clustering procedure) and the new M segments to cluster.]

25